cases where, unintentionally, the ILEC provides CLECs with a level of performance that is better than the performance it provides to itself.[12] Thus, the Commission's rules themselves argue for a one-tailed test.

22. It should also be recognized that Type II errors are as real as Type I errors. Thus, there may be cases in which the ILEC is not in fact providing equal service to CLECs, but purely by chance the statistical test fails to reject the parity hypothesis. Thus, it is necessary to strike a balance between the two types of errors. If we choose to make the Type I error small, then the Type II error will be large; and conversely. AT&T proposes to set the Type I error at no more than the conventional level of 5%. This controls the frequency of false alarms to be at most 5% while making the probability of Type II errors small for violations that are of substantial size. Using a one-tailed test for Type I error at about the 5% level thus strikes a reasonable balance.[13]

---

[12] I am also informed that CLECs are not entitled to demand performance better than the ILEC provides to itself. Thus, there is no reason to believe that ILECs would intentionally provide their competitors with a higher grade of service than they provide to themselves and their retail customers.

[13] For general information supporting the 5% level, see AT&T Statistical Ex Parte, pp. B-1-B-2.

C. Probability Distribution Should Be Based On The T Distribution Or A Permutation Distribution Analysis.

23. For moderate or large sample sizes, it is appropriate to use the Student t (or "t") distribution to determine the critical value for the test. Use of this distribution, which is readily available in table form, is simple and straightforward and will produce statistically reliable results.

24. The published tables of critical values, using the t-distribution, are based on the assumption that the two populations (of ILEC and CLEC measurements) are exactly Normal. In practice, we will not have Normal distributions, and so these critical values are only approximations. There has been much debate as to the minimum sample sizes for which the tabulated values become acceptable approximations: numbers such as 10 or 30 have been suggested. But this must depend on the shape of the probability density function[14] of the populations, because there exist populations for which the approximation will never be adequate, even for very large sample sizes. In advance of reviewing the actual data, it is impossible to say for what size samples the tabulated values will be acceptable. Nevertheless, assuming that very large values of the observations do not occur and the populations have approximately symmetrical probability

---

[14] See the graph in ¶14 of these comments for an example of a probability density function.

density functions, I would guess that the tabulated values would be acceptable, provided that both the ILEC and CLEC samples have at least 10 members. Thus, the issue of sample size should not generally be a problem.

25. There is an alternative method for developing the probability distribution of the test statistic that can be used with smaller sample sizes.[15] Under this method, called the permutation distribution, the probability distribution is generated through use of the actual sample results, rather than a preexisting table. Given two samples, X's and Y's from ILEC and CLEC respectively, we combine these into one pool and then divide this into two sets X* and Y* in all possible ways. For each way, we calculate the corresponding z-score, say z*. This gives us a distribution of z* values, each of which is equally likely under the null hypothesis that the ILEC is treating customers impartially. Given the desired Type I error rate, we can read off the appropriate critical value and compare this with the observed value.

26. For example, if the data are

3 ILEC observations: X=1, X=2, X=4
2 CLEC observations: Y=3 and Y=5

---

[15] This method will provide reliable results for any sample size, but the use of the t-distribution and the associated table is simpler for all but very small sample sizes.

then the pooled set is (1,2,3,4,5) and there are 10 ways we can assign these five observations to the ILEC and CLEC samples. We get 10 values of z:

-2.74 -1.20 -0.60 -0.44 0.00 0.00 0.44 0.60 1.20 2.74

and the 5% critical value is 2.74. The actual observed value is 1.20, and so is judged to be not significant (i.e., we accept the null hypothesis).

27. This test procedure is valid irrespective of the form of the population distribution, since it depends only on the assumption that each possible permutation is equally likely under the null hypothesis.[16] The method can be used whenever the sample sizes are large enough to make the test statistic well defined, in the present case even for m=2, n=1.

28. The permutation distribution would be developed through the use of a computer program that would enumerate the samples necessary to generate the distribution. I wrote a program to perform this function in a commercially available program language called S Plus in one-half hour. Thus, I believe that a suitable program could be developed

[16] See, e.g., Cox and Hinkley, Theoretical Statistics (1974) (paperback edition Chapman and Hall, 1979), pp. 182-184; H. Scheffe, The Analysis of Variance (1959) John Wiley & Sons, Section 9.3; P. Good, Permutation Tests (1994) Springer.

promptly for use by the entire ILEC industry at minimal cost[17].

29. A resource issue relating to the use of the permutation distribution is the time needed to generate results. Unless the sample sizes are very small, the number of permutations to be generated is extremely large.[18] In order to deal with this problem, it would be reasonable to use a random sample of possible permutations to approximate the distribution. For example, if the number of possible permutations in a particular case exceeds 1000, the program could be designed to approximate the permutation probability distribution by randomly selecting 1000 permutations and constructing the distribution from those data. Because computers can perform calculations such as this with remarkable speed, the distribution for any measurement category could be ascertained within a few seconds.[19]

---

[17] The Cytel Software Corporation of 675 Massachusetts Avenue, Cambridge, MA, markets a product called StatXact which has the capability of performing permutation tests.

[18] If m=10, n=5, there are 3003 permutations; if m=20, n=10, there are over 30 million.

[19] The Notice (Appx. B, n.5) raises another interesting possibility for a statistical analysis of individual performance measurements, i.e., comparing the proportions of two samples that exceed some fixed value. AT&T is studying a variation of this concept, in which the fixed value is not specified in advance, but is determined from the ILEC sample itself. We use the upper 90% quantile of the ILEC sample to determine the level of service that the ILEC is providing for 90% of its customers and then measure what percentage of CLEC customers receive at least that level of service. The

D. ILECs' Compliance With Their Nondiscrimination Obligations Should Be Based On An Aggregate Assessment Of Parity.

30. One of the key concepts in the AT&T Statistical Ex Parte is that it is also appropriate to use statistical analysis to review the aggregate results of an ILEC's performance to determine whether it is in compliance with its nondiscrimination obligations. If we apply a large number, several hundred perhaps, of tests of individual performance measurement comparisons, each test having a Type I error rate of 5%, then we would expect, on average, about 5% of these tests to indicate non-compliance even when the ILEC is actually fully in compliance. Thus the fact that this many tests indicate non-compliance does not give conclusive evidence that the ILEC is not in compliance with its Section 251 nondiscrimination obligations. The number of tests that erroneously indicate non-parity will vary randomly about this average number. We need to derive some

---

"parity" hypothesis is rejected if the fraction of CLEC customers receiving that level of service is much smaller than the percentage of ILEC customers receiving such service. (For example, if the ILEC completes repairs on a specific service for 90% of its customers within 48 hours, parity is not achieved if the ILEC complete repairs for much less than 90% of CLEC customers within that amount of time.) This test procedure is non-parametric, <u>i.e.</u>, it does not require any assumptions beyond the basic one that under the null hypothesis CLECs receive equal treatment to the ILEC. This methodology only applies, however, to the review of individual performance tests. It does not address the need to develop a method to review ILEC performance in the aggregate.

threshold number of failed parity tests such that if more than this number are observed to fail, then non-compliance can be deduced. This threshold number of tests must be determined in such a way as to control the probability of an overall, or aggregate, Type I error at 5%. Furthermore, I also recommend that any review of an ILEC's compliance with its nondiscrimination obligation should be based on two dimensions of statistical comparisons, both of which must be satisfied.[20] The two dimensions of statistical comparisons are

(a) the number of tests that fail in any monthly period must not be too large, and

(b) the number of tests that fail for three consecutive months must not be too large.

Here, "too large" must be determined by consideration of the total number of individual tests and the desired overall Type I error rate.

31. For the first dimension, we must determine how many of the individual measurements subjected to the above comparison tests need to demonstrate non-parity before an ILEC may be found to be in overall violation of its

---

[20] The AT&T Statistical Ex Parte suggested that a third dimension also be considered, namely imposing a bound on the number of individual tests that exhibit extreme violations. I now judge that imposing this additional constraint does not provide much additional power for detecting extreme violations, and in fact reduces the chance of detecting some more moderate violations.

statutory duty. Suppose we have made N individual tests, each having a 5% Type I error rate, and have found that K_1 of them indicate non-compliance. If K_1 is approximately .05 times N, we have no conclusive evidence of overall non-compliance. Under the assumption that the ILEC is in compliance, we can determine a number k_1 such that the probability that K_1 exceeds k_1 is 5%.[21]

32. The second dimension, i.e., the number of measurements failing the test repeatedly, is necessary to assure that the ILEC failures are indeed random. Without this dimension, the ILEC might be able to "game" the process and produce repeatedly discriminatory results on measures that are critical to one or more competitors. Thus, for this dimension, we must determine how many individual measurements in an ILEC report may be allowed to fail the parity test in three successive months before finding that the ILEC has failed to provide parity.

33. Suppose we have made N individual tests for each of three months, each test having a Type I error of 5%. Let K_2 be the number of tests that have failed in all three months. The probability that any individual test fails in

---

[21] This computation assumes that under the null hypothesis, the number K_1 has a binomial distribution with exponent N, i.e., it is as though we had tossed N coins, each with a probability of coming down "heads", and have counted how many "heads" appear. Then we claim non-compliance if K_1 exceeds k_1.

all three months, assuming that the ILEC is in compliance with its nondiscrimination obligation, is $(.05)^3$, or 1/8000. Thus the expected number that fail in all three months, assuming compliance, is N/8000. Given that the number of monthly tests will be well below 8000, noncompliance should be found if K_2 is not zero. In other words, the allowed number of three-time-failing tests is $k_2=0$.

34. If we apply both of these overall procedures simultaneously, the actual overall Type I error rate is a function of three things: the Type I error rates of the individual tests, which I call α_1, the number k_1 of allowed individual failures, and the number k_2 of allowed three-time failures. These three numbers can be determined so that the Type I error rate of the overall procedure is exactly 5% (or whatever other value is required). Details of this computation are given in Exhibit 1.

II. BellSouth's Proposed Methodology Is Unsuited To Measure Parity And Should Be Rejected.

35. The Notice (Appx. B, ¶ 7) also solicits comments on the methodology proposed by BellSouth, which is based on the use of statistical process control. This approach is not suitable to measure parity between ILECs and CLECs and should be rejected.

36. BellSouth has proposed three kinds of control charts. In the first, described in the Notice (Appx. B,

¶ 6), BellSouth maintains its own monthly results (presumably for each type of measurement) on a control chart. Three-sigma limits are established by reference to BellSouth's historical record. Then, each month, results for the CLEC are plotted on the same chart, and parity is claimed if these values do not fall outside the limits.

37. A second proposal appears in BellSouth's Tennessee Section 271 proceeding (see memo from David Laney to William Stacy, attached to the rebuttal testimony of William N. Stacy, TRA Docket 97-00309, Exhibit WNSPM-2). Here the proposal is to plot values of the variable DIFF=(CLEC value - ILEC value) on a control chart, with limits set at +/- 2.66 times the average moving range of size two.

38. A third proposal also appears in the same document from BellSouth's Tennessee Section 271 proceeding. Here it is proposed to compute z-scores, but using the process standard deviation in the denominator rather than the within-month ILEC sample variance as AT&T recommends. This process standard deviation is the average moving range (presumably of size two) divided by 1.128.

39. Each of these proposals has serious deficiencies, the most serious being that statistical process control is not designed to measure differences in parity. Rather, this technique is used to measure stability in performance. Stability of ILEC processes is of course an important

concept, because the overall reliability of the systems used to serve CLECs is essential to determining whether an ILEC has met its duties under Section 251 of the Act. However, it is irrelevant in determining whether an ILEC's performance for itself is at parity with the performance it provides to others, i.e., CLECs. The ILEC's performance could be stable, with parity not provided, or unstable with parity being provided. Stability and parity are distinctly different concepts.

40. Another shortcoming of each of the three BellSouth proposals is that no allowance is made for the fact that the number of observations that contribute to each average may change from month to month. This makes the use of moving ranges invalid measurements of variability. Also, the number of observations in the CLEC sample is very unlikely to equal the number in the ILEC sample. Thus the ILEC and CLEC averages will not have the same variances, even assuming parity, and so should not be compared to the same control limits, as the first proposal suggests.

41. If control limits for the quantity DIFF were to be set using the process variability of this quantity, as in the second and third proposals, some consistent violations of parity could completely avoid detection. Namely, if for any reason the CLEC measurements were consistently more variable than the ILEC measurements (which would imply that

many CLEC customers were getting poorer service), then this variability would be included in setting the control limits, and lack of parity would not be detected.

42. Further, use of separate control charts for each of the many types of measurement leaves open the question of how an overall judgement of compliance should be arrived at. BellSouth has not addressed this issue.

Conclusion

43. In summary, my testimony shows that AT&T's proposed methodology satisfies the Commission's desire to assure that reported differences in ILEC performance are statistically meaningful.

44. With respect to individual tests of ILEC performance, there are three key components in developing an appropriate statistical methodology. First, the modified z-statistic proposed by LCUG provides an appropriate test statistic to determine whether there are significant differences in the mean and the variance of an ILEC's performance for itself and for CLECs. Second, a one-tailed test with Type I error at about the 5% level strikes a fair balance between the need to account for both Type I and Type II errors. Third, the t-distribution provides a useful basis for calculating the critical value for individual tests of ILEC performance. Moreover, in those few cases

where the size of the ILEC sample is small, use of the permutation distribution will provide valid results.

45. It is also appropriate to aggregate the results of individual tests to determine whether the ILEC is in overall compliance with its duty to provide nondiscriminatory treatment to CLECs. This should be done through the use of a two-part analysis that sets limits on the number of individual tests that fail to demonstrate parity in any given month and on the number of individual tests that fail in three consecutive months. These limits can be determined in such a way that the overall Type I error is held at 5%.

46. Finally, the methodology suggested by BellSouth is not designed to measure parity of performance between two different populations. Thus, it should not be used to determine whether ILECs have met their legal duty to provide CLECs with parity service.

Colin L. Mallows

Colin L. Mallows

Sworn to before me this
29th day of May, 1998

Patricia A. Perhae
Notary Public

My Commission expires 4/8/2002

Exhibit 1

Statistical Definition of the Compliance Rule for ILEC Parity

The number k_1 of allowed individual violations, and the Type I error of each of the individual tests[22], α_1, are determined so that the probability of falsely claiming overall violation is controlled at a known level[23], which we call α.

Suppose we are aggregating N individual tests. Let K_1 be the number of these tests that indicate violations this month, and let K_2 be the number of tests that have shown violations in each of the past three months. Our proposed procedure is to claim overall violation if either (i) K_1 exceeds some number k_1, or (ii) K_2 exceeds zero. We show how k_1 and the type I error α_1 of each individual test can be determined so that the Type I error of the overall procedure is held at some desired level α.

To determine k_1 and α_1 when we know N, (the number of tests to be aggregated), and α, we proceed as follows. Throughout this calculation, we are assuming that the ILEC is fully in compliance, so that for each individual test the probability of (falsely) indicating non-parity is α_1.

a) Choose a tentative value for α_1. We start with $\alpha_1 = \alpha$. This value of α_1 will be adjusted (downwards) later.

b) Determine k_1 to be the largest number such that the probability that the overall procedure indicates violation[24] (is greater than α.

c) Decrease α_1 until the probability of overall violation using the value of k_1 that was determined in step b), is exactly α.

---

22 Also referred to as the size of the individual test.

23 Also referred to as the size of the overall aggregated test.

24 This probability is: $1 - (1 - \alpha_1^3)^N * P(k, N, p)$ where P(,,) is the cumulative probability of the binomial distribution. That is, P(k, N, p) is the probability that the number of false parity test failures is <= k when the probability of an individual false parity test failure is p, and where $p = (\alpha_1 - \alpha_1^3)/(1 - \alpha_1^3)$.

The resulting value of α_1 (and the corresponding critical value on the z-score scale) is to be used in each of the individual tests. Then non-compliance is indicated if any series fails the test in three successive months, or if more than k_1 fail in any single month.

The following table provides an example of how k_1 is determined for the values N = 100 and α = 5%. As shown, the value of k_1 = 8 is the largest value of k that corresponds to a probability of no less than 5% of being exceeded. In this case, the probability of claiming an overall violation is 7.40%.

Table 1

Determination of k_1 for N=100, α= 5%

k	Prob{K_1>k, K_2>0} =1 - (1 - $\alpha_1^3)^N$ * P(k, N, p)	
5	38.95%	
6	24.17%	
7	13.76%	
8	7.40%	← select this k for k_1
9	3.99%	
10	2.36%	

The next step is to iteratively decrease α_1 and recompute the overall probability of violation, with k_1 held at 8, until we arrive at a value for α_1 for which this probability is .05. In this case, that value of α_1 is .04601.

Now we can use the t-tables (or permutation distribution calculations) to determine the appropriate critical values for each individual test. The following Table 2 provides k_1, α_1, and critical values (assuming large sample sizes for each test) for α = .05 and a number of values of N.

Table 2

Determination of k_1 and α_1 for a range of N where k_1 satisfies $1 - (1 - \alpha_1^3)^N * P(k_1, N, p) = .05$

N	k_1	k_1 as a % of N	α_1	Critical Value (c)
70	6	8.57%	.0465	1.6803
80	6	7.50%	.0408	1.7411
90	7	7.78%	.0437	1.7096
100	8	8.00%	.0460	1.6849
120	9	7.50%	.0442	1.7038
140	10	7.14%	.0430	1.7170
160	12	7.50%	.0462	1.6825
180	13	7.22%	.0452	1.6937
200	14	7.00%	.0443	1.7026
250	17	6.80%	.0441	1.7046
300	20	6.67%	.0440	1.7060
400	26	6.50%	.0437	1.7095
500	32	6.40%	.0431	1.7155
600	38	6.33%	.0423	1.7247
700	44	6.29%	.0412	1.7374
800	49	6.13%	.0397	1.7543
900	55	6.11%	.0384	1.7696
1000	60	6.00%	.0371	1.7851

Table 1
Regulatory Summary Report
Performance Measurement Compliance Summary Report
Aggregate CLEC Summary - Report Period: May 1998

	Measurements Failed This Period(α=.0440, c=1.7060)	Measurements Failed In Three Consecutive Months
Total Comparisons of Measurement/Reporting Dimension Combinations for the specified time period	300	300
Threshold	20	0
Number of Failed Comparisons	26	0
Compliance Assessment	ILEC Performance Measurements Not In Compliance	

Table 2

Regulatory Detail Exception Report

CLEC Aggregate Report - Report Period: May 1998

Number Measurements Failed This Month

ID	Description	Reporting Dimension	Number Of CLECs Failing Test	ILEC			CLEC Aggregate		Test
				n	mean	variance	n	mean	z
OP-1	Average Completion Interval(days)	/Resold Bus POTS/ New service Installation/	5	4000	2.8	70.56	500	3.9	2.66
OP-1	Average Completion Interval(days)	/Resold Res POTS/ New service Installation/	5	8000	1.1	19.36	500	1.7	2.85
OP-3	Percent Order Accuracy	/Resold Bus POTS/	7	4000	95%	0.05	500	93%	1.75
OP-9	Mean Held Order Interval (days)	/Unbundled DSO Loop/No Facilities/	3	4000	10	900.00	300	15.2	2.87
MR-1	Mean Time To Restore (hrs)	/Unbundled DSO Loop/Loop, Access Line/	3	3000000	12	4356.00	3500	14.8	2.53
MR-2	Repeat Trouble Rate	/UNE-P/loop- access line/	3	3000000	9%	0.08	3500	12%	5.63
OS/DA-1	Mean Time to Answer (secs)	/Operator Services/	3	45000	10	900.00	450	13.1	2.15

Table 3

Individual CLEC Complete Detail Report

Report for CLEC A- Report Period: May 1998

(Report includes Detail CLEC Aggregate and Detail ILEC Results)

ID	Description	Reporting Dimension	ILEC			CLEC Aggregate		Test	CLEC A		Test
			n	Mean	Variance	n	Mean	z	n	Mean	z
PO-1	Average Response Interval	/Due Date reservation/	nn	x.xx	yyy.y	nn	x.xx	yyy.y	nn	x.xx	yyy.y
...	...	...	...	...	...	...	...	...	...	...	...
OP-1	Average Completion Interval	/Resold Residence POTS/New Service Installations/	nn	x.xx	yyy.y	nn	x.xx	yyy.y	nn	x.xx	yyy.y
...	...	...	...	...	...	...	...	...	...	...	...
MR-1	Mean Time To Restore	/ Resold Residence POTS /Out of Service No Dispatch/	nn	x.xx	yyy.y	nn	x.xx	yyy.y	nn	x.xx	yyy.y
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
IUE-2	Timeliness of Element Performance	/Unbundled Loop/	nn	x.xx	yyy.y	nn	x.xx	yyy.y			
...	...	...	...	...	...	...	...	...	...	...	...